

# Machine Learning (MC 321)

## Lab Assignment 6

March 3, 2025

### Question 1

Perform binary classification using logistic regression on the data in file “Social Network Ads.csv”, which is a categorical dataset to determine whether a user purchased a product or not by using three features.

1. Visualize the data by 3D plotting features using different colors for label 0 and 1.
2. Implement the logistic function and learn the model parameters using:
  - Stochastic gradient descent method.
  - Gradient descent with momentum and Regularization.
  - Newton’s method (iteratively reweighted least squares method) and Regularization. Does the IRLS method converge in fewer iterations than GD?
  - Do a performance comparison between gradient descent with momentum and Newton’s method.
3. Plot the training data, test data, and decision boundary learnt by logistic regression in the same figure. (It should be a straight line showing the boundary separating the region where  $h_w(x) > 0.5$  from the region where  $h_w(x) \leq 0.5$  ( $h_w(x^{(i)}) = \sigma(w^T x^{(i)})$ )) for the above three results. Use 90% data points from each set for training and the remaining 10% for testing the accuracy of classification.

### Question 2

Consider the Iris flower dataset that contains a set of 150 samples, which consists of 50 samples from each of three species of Iris: setosa (label 0), versicolor (label 1), and virginica (label 2). Each sample was measured in four features: sepal length, sepal width, petal length, and petal width.

#### Data Preparation and Visualization:

- Split the dataset into a balanced (with respect to the labels) training and test set, containing respectively 80% and 20% of the dataset.
- Visualize the data by 3D plotting features using different colors for label 0,1 and 2.
- Implement the logistic function and learn the model parameters using:
  - Stochastic gradient descent method.
  - Gradient descent with momentum and Regularization.
  - Newton’s method (iteratively reweighted least squares method) and Regularization. Does the IRLS method converge in fewer iterations than GD?
  - Do a performance comparison between gradient descent with momentum and Newton’s method.
- Plot the training data, test data, and decision boundary learnt by logistic regression in the same figure. (It should be a straight line showing the boundary separating the region where  $h_w(x) > 0.5$  from the region where  $h_w(x) \leq 0.5$  ( $h_w(x^{(i)}) = \sigma(w^T x^{(i)})$ )) for the above three results. Use 90% data points from each set for training and the remaining 10% for testing the accuracy of classification.